Aufgabe 1

Bei einer Polizeikontrolle werden die erfassten Fahrzeuge unter anderem auf die Merkmale „Fahrzeugtyp" (Motorrad(MR), PKW oder LKW) und „Geschlecht des Fahrzeugführers" (Männlich(M) oder Weiblich(F)) untersucht.

1. Um welchen Merkmalstyp handelt es sich beim Merkmal „Fahrzeugtyp"? Welche Skalierung ist für diesen Merkmalstyp sinnvoll?

2. Vervollständigen Sie die Tabelle der relativen Häufigkeiten ($f(\cdot)$).

	MR	PKW	LKW	Σ
M	0.11		0.04	0.7
F			0.01	
Σ	0.15	0.8		1

3. Interpretieren Sie die relative Häufigkeit $f(M|MR) \approx 0.73$.

4. Nennen und bestimmen Sie ein geeignetes Lagemaß für das Merkmal „Fahrzeugtyp".

5. Nehmen Sie auf Grundlage der Tabelle aus Teilaufgabe 2. zu folgenden Aussagen Stellung und begründen Sie diese:

   a) Mehr als 30% der kontrollierten männlichen Fahrer fahren Motorrad.

   b) Weniger als 60% der kontrollierten Fahrer sind männlich und fahren PKW oder LKW.

   c) Der Anteil der Frauen unter den kontrollierten LKW-Fahrern ist größer als der Anteil der LKW-Fahrer unter den kontrollierten Männern.

Laut Kraftfahrtbundesamt[1] waren in Deutschland 2014 71% der Motorradfahrer, 50% der PKW-Fahrer und 64% der LKW-Fahrer <u>männlich</u>.
Unter der Annahme, dass es nur die drei Fahrzeugtypen „Motorrad" (MR), „PKW" und „LKW" gibt, sind außerdem folgende Wahrscheinlichkeiten bekannt:

$P(MR) = 8\%$ $\qquad$ $P(PKW) = 87\%$

[1] http://www.kba.de/DE/Statistik/Kraftfahrer/Fahrerlaubnisse/Fahrerlaubnisbestand/fahrerlaubnisbestand_node.html

6. Berechnen Sie die Wahrscheinlichkeit, dass ein Fahrer männlich ist, d.h. $P(M)$.

Seien

X_M : „Anzahl der männlichen Fahrer"

X_F : „Anzahl der weiblichen Fahrer"

sowie

Y_i : „Anzahl der Fahrer eines Fahrzeugs vom Typ i"

mit $i \in \{$„PKW", „MR", „LKW"$\}$

7. Berechnen Sie die Wahrscheinlichkeit, dass

   a) unter 2 Motorrad-Fahrern und 3 LKW-Fahrern mindestens ein Fahrer männlich ist, d.h. $P(X_M \geq 1 | Y_{MR} = 2, Y_{LKW} = 3)$,

   b) unter 5 PKW-Fahrern höchstens zwei Fahrer weiblich sind, d.h. $P(X_F \leq 2 | Y_{PKW} = 5)$.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Es wurde die Körpergröße von 8 zufällig ausgewählten Ehepaaren erhoben und in der folgenden Tabelle aufgeschrieben.

Ehepaar i	Körpergröße m_i des Ehemanns (in cm)	Körpergröße f_i der Ehefrau (in cm)
1	176	174
2	181	160
3	171	170
4	168	173
5		165
6	181	168
7	182	180
8	169	171

Mit den vorliegenden Daten soll die Vermutung getestet werden, dass Ehemänner im Mittel größer als ihre Ehefrauen sind. Fassen Sie die Einträge in obiger Tabelle als Realisationen einer Zufallsstichprobe mit den Zufallsvariablen

$$M_i : \text{„Körpergröße des } i\text{-ten Ehemanns"} \quad \text{und} \quad F_i : \text{„Körpergröße der } i\text{-ten Ehefrau"}$$

auf.

Realisierte Statistiken, die Sie bitte im Folgenden verwenden, sind $\overline{m}_8 = 176.5000$, $\overline{f}_8 = 170.1250$, $s_M^2 = 40.8571$, $s_F^2 = 36.4107$, $s_{MF} = -10.2143$.

1. Bestimmen Sie die Größe des fünften Ehemanns (m_5) in cm.

2. Berechnen Sie den Stichprobenkorrelationskoeffizienten r_{MF}.

3. Die Stichprobenvarianz von $D = M - F$ wird gegeben durch

$$s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left((m_i - f_i) - (\overline{m}_n - \overline{f}_n) \right)^2 .$$

Zeigen Sie folgende Gleichheit:

$$s_D^2 = \frac{\sum_{i=1}^{n} (m_i - \overline{m}_n)^2}{n-1} + \frac{\sum_{i=1}^{n} (f_i - \overline{f}_n)^2}{n-1} - 2\frac{\sum_{i=1}^{n} (m_i - \overline{m}_n)(f_i - \overline{f}_n)}{n-1}.$$

4. Führen Sie einen Mittelwertdifferenzentest für verbundene normalverteilte Stichproben zum Niveau $1 - \alpha = 0.9$ durch. Das Hypothesenpaar lautet:

$$H_0 : \ \mu_M \leq \mu_F \quad \text{gegen} \quad H_1 : \ \mu_M > \mu_F$$

a) Berechnen Sie für $D = M - F$ die Stichprobenvarianz s_D^2.

Verwenden Sie im Folgenden $s_D^2 = 100$, auch wenn Sie in Teilaufgabe 4a) einen anderen Wert herausbekommen haben.

b) Berechnen Sie die realisierte Prüfgröße t_n.

c) Geben Sie die realisierte Untergrenze des einseitigen Konfidenzintervalls an, das dem Test zugehörig ist.

Verwenden Sie im Folgenden die Untergrenze 1.23**, auch wenn Sie in Teilaufgabe 4c) einen anderen Wert herausbekommen haben**.

d) Treffen Sie die Testentscheidung basierend auf dem Konfidenzintervall.

e) Wie lautet die Testentscheidung, wenn das Hypothesenpaar zu

$$H_0: \ \mu_M \leq \mu_F + 1 \quad \text{gegen} \quad H_1: \ \mu_M > \mu_F + 1$$

geändert wird?

f) Wird das Niveau geändert, ändert sich die Länge eines Konfidenzintervalls. Wird es länger oder kürzer, wenn α vergrößert wird?

5. Der Schätzwert für das dritte standardisierte Moment der Verteilung von D beträgt $\mu_3 = 0.4511$.

a) Interpretieren Sie diesen Wert.

b) Wenn M und F bivariat normalverteilt sind, ist $D = M - F$ ebenfalls normalverteilt. Welches Schiefeverhalten weisen Normalverteilungen auf?

6. Es wird ein weiteres Vorgehen vorgeschlagen: Stammen die Größen von Männer und Frauen alle aus derselben Grundgesamtheit und finden Paare unabhängig von der Körpergröße zueinander, dann sind die Zufallsvariablen

$$G_i = \begin{cases} 1 & \text{falls der } i\text{-te Ehemanns größer ist, als seine Frau} \\ 0 & \text{falls der } i\text{-te Ehemanns nicht größer ist, als seine Frau} \end{cases}$$

jeweils Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p = 0.5$. Darüber hinaus sind diese Zufallsvariablen voneinander unabhängig.

a) Geben Sie die vollständig spezifizierte Verteilung der Zufallsvariablen $Z = \sum_{i=1}^{8} G_i$ an.

b) Berechnen Sie die Realisation von Z. Gehen Sie davon aus, dass $m_5 > f_5$, **unabhängig von den Berechnungen in Teilaufgabe 1**.

7. a) Eine faire Münze wird 8 Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass diese genau 2 Mal Kopf zeigt?

b) Eine faire Münze wird 8 Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass diese höchstens 2 Mal Kopf zeigt?

Schmierpapier zu Aufgabe 2

Aufgabe 3

Laut offizieller FIFA-Statistik erzielte die deutsche Fußballnationalmannschaft in den letzten 10 Spielen durchschnittlich 2.2 Tore pro Spiel, wobei im Mittel 0.5 Tore in der ersten Hälfte des Spiels erzielt wurden. Im Folgenden seien

X_i: „Anzahl der in Halbzeit i geschossenen Tore”

mit $i = 1, 2$. Weiter sei angenommen, dass die X_i unabhängig poissonverteilt mit Parameter λ_i für $i = 1, 2$ ist.
Dann ist

Z: „Anzahl der geschossenen Tore in einem Spiel”

mit $Z = X_1 + X_2$ ebenfalls poissonverteilt mit Parameter $\lambda_Z = \lambda_1 + \lambda_2$.

1. Geben Sie die geschätzten Parameter λ_1, λ_2 sowie λ_Z auf Grundlage der letzten 10 Spiele an.

Verwenden Sie für alle nachfolgenden Teilaufgaben $\lambda_1 = 1.5$, $\lambda_2 = 1$ sowie $\lambda_Z = 2.5$

2. Berechnen Sie die Wahrscheinlichkeit, dass Deutschland in der zweiten Halbzeit des nächsten Länderspiels kein Tor erzielt?

3. a) Wie hoch ist die Wahrscheinlichkeit, dass Deutschland im nächsten Spiel mehr als zwei Tore erzielt?

b) Wie hoch ist die Wahrscheinlichkeit, dass Deutschland in der ersten Halbzeit mindestens ein Tor, aber nicht mehr als vier Tore schießt?

4.) Mittels eines approximativen Tests soll

$$H_0 : \lambda_Z = 3 \text{ gegen } H_A : \lambda_Z \neq 3$$

auf einem Signifikanzniveau von 10% getestet werden, d.h. die Teststatistik $T_n = \sqrt{n}\frac{\bar{Z}_n - \lambda_0}{\sqrt{\lambda_0}}$ wird als approximativ standardnormalverteilt angenommen.

a) Berechnen Sie die zugehörige Teststatistik.
Beachten Sie, dass ein konsistenter Schätzer für den Parameter λ_Z durch $\overline{Z}_n$ gegeben ist. Verwenden Sie im Folgenden $\overline{z}_{10} = 2.5$

b) Geben Sie den kritischen Bereich des Tests sowie Ihre Testentscheidung an.

c) Berechnen Sie für den approximativen Test den p-Wert. Gehen Sie hierbei von einem realisierten Wert der Teststatistik von 1.48 aus (**auch wenn Sie in Teilaufgabe 4a) einen anderen Wert herausbekommen haben**).

Sollten Sie aus Teilaufgabe 4c) kein Ergebnis vorliegen haben, so verwenden Sie im Folgenden einen p-Wert von 0.13

d) Fällen Sie Ihre Testentscheidung auf Basis des p-Werts aus Teilaufgabe 4c).

e) Geben Sie das größtmögliche Signifikanzniveau an, für welches Sie in Teilaufgabe 4d) die Nullhypothese nicht ablehnen können?

5. Die deutsche Fußballnationalmannschaft wurde in der Vergangenheit oft für ihre schlechte Chancenauswertung kritisiert. Die geometrische Verteilung kann z.B. dazu herangezogen werden, um die Anzahl der vergebenen Torchancen bis zum ersten Tor zu modellieren.

   Die Wahrscheinlichkeitsmassenfunktion f_X, und die kumulative Verteilungsfunktion F_X der geometrischen Verteilung lauten:

$$f_X(x;p) = p \cdot (1-p)^x \ , \tag{1}$$

$$F_X(x;p) = 1 - (1-p)^x \ , \tag{2}$$

   für $x = 0, 1, 2...$, wobei der Parameter p die Wahrscheinlichkeit eines Erfolgs misst.

a) Zeigen Sie, dass die Loglikelihoodfunktion der geometrischen Verteilung folgende Form besitzt:

$$LL(p; x_1, ..., x_n) = n \cdot ln(p) + ln(1-p) \cdot \sum_{i=0}^{n} x_i.$$

b) Leiten Sie den Maximum-Likelihood Schätzer für den Parameter p der geometrischen Verteilung her und geben Sie den Schätzwert an.

Ihnen sei nun bekannt, dass die deutsche Mannschaft in der Vergangenheit durchschnittlich 5.5 Torversuche pro Spiel benötigte bis zum ersten Mal ein Tor erzielt wurde. *(Sollte Ihnen aus Teilaufgabe 5b) kein Ergebnis vorliegen, verwenden Sie nachfolgend* $\widehat{p}_{ML} = 0.17$*)*

c) Wie hoch ist dann die geschätzte Wahrscheinlichkeit, dass beim vierten Torschuss das erste Tor erzielt wird?

d) Wie hoch ist die geschätzte Wahrscheinlichkeit, dass mindestens zwei aber weniger als sieben Torschüsse bis zum ersten Tor benötigt werden?

Schmierpapier zu Aufgabe 3

Aufgabe 4

Die beiden Fußballfans A und B haben nun schon drei Lokalitäten zum WM-Spiel-Schauen ausgetestet. Da es in der Woche vor den ersten Spielen viel geregnet hatte, waren viele Stechmücken unterwegs. A und B wollen Untersuchungen bezüglich der Mückenstiche, die sie an den drei Lokalitäten (X, Y oder Z) abbekommen haben durchführen.

1. Welchen Merkmalstyp hat das Merkmal *Größe der Mückenstiche* mit den Ausprägungen *klein* (2 Beobachtungen), *mittel* (2 Beobachtungen) und *groß* (8 Beobachtungen)?

2. Geben Sie das Lagemaß für das Merkmal *Größe der Mückenstiche* aus Teilaufgabe 1. an, welches zu dessen Skalentyp passt.

Im Folgenden soll nun untersucht werden, ob die Wahl der Lokalität einen Einfluss auf die Anzahl der Stiche bei A und B hat. Dazu haben beide die erhaltenen Stiche in jeder Lokalität gezählt und folgende Tabelle erstellt:

		Lokalität			
		X	Y	Z	Σ
Fußballfan	A	2	0	3	5
	B	4	2	1	7
	Σ	6	2	4	12

3. Berechnen Sie die Anzahl von Beobachtungen, die sich bei Unabhängigkeit der Merkmale *Lokalität* und *Fußballfan* ergeben würden und tragen Sie diese in folgende Tabelle ein

		Lokalität			
		X	Y	Z	Σ
Fußballfan	A				5
	B				7
	Σ	6	2	4	12

Gehen Sie unabhängig von den bisherigen Ergebnissen ab sofort von folgender Indifferenztabelle aus:

		Lokalität			
		X	Y	Z	Σ
Fußballfan	A	$n_{11}^* = 2$	$n_{12}^* = 1$	$n_{13}^* = 2$	5
	B	$n_{21}^* = 3$	$n_{22}^* = 2$	$n_{23}^* = 2$	7
	Σ	6	2	4	12

4. Berechnen Sie die mittlere quadratische Kontingenz ϕ^2 der beiden Merkmale *Lokalität* und *Fußballfan*.

5. Geben Sie allgemein für Kontingenztabellen der vorliegenden Dimensionen die kleinste obere und die größte untere Schranke für ϕ^2 an.

Gehen Sie unabhängig von bisherigen Ergebnissen ab sofort von $\phi^2 = 0.5367$ aus.

6. Berechnen Sie Cramers V!

7. Geben Sie allgemein die kleinste obere und die größte untere Schranke für Cramers V an.

8. Welchen Wert würden die Größen ϕ^2 und V jeweils bei Unabhängigkeit der beiden Merkmale *Lokalität* und *Fußballfan* annehmen?

Eigentlich ist die Anzahl der Beobachtungen zu gering um asymptotische Tests durchführen zu können. Gehen Sie im Folgenden dennoch davon aus, dass für die obige Kontingenztabelle die Größe $n\phi^2$ näherungsweise χ^2 -verteilt ist mit 2 Freiheitsgraden.

9. Wie lautet der kritische Bereich für $n\phi^2$, der zum Hypothesenpaar

   H_0 : Merkmale Lokalität und Fußballfan sind stochastisch unabhängig

   H_1 : Merkmale Lokalität und Fußballfan sind nicht stochastisch unabhängig

   mit Irrtumswahrscheinlichkeit $\alpha = 10\%$ gehört?

10. Berechnen Sie den p-Wert für eine Realisation der Teststatistik aus Teilaufgabe 9. von $\phi^2 = 0.5367$.

Gehen Sie unabhängig von den bisherigen Ergebnissen ab sofort von einem p-Wert von 0.12 aus.

11. Treffen Sie eine Testentscheidung anhand des p-Werts.

Schmierpapier zu Aufgabe 4

Lösung Klausur SS14 (7.5ECTS)

Aufgabe 1

1. qualitatives Merkmal (ungeordnet kategorial)✓, Nominalskala✓ (0.5P+0.5P)

2. ✓✓+ Rechenweg ✓ (1P+0.5P)

	MR	PKW	LKW	Σ
M	0.11	0.55	0.04	0.7
F	0.04	0.25	0.01	0.3
Σ	0.15	0.8	0.05	1

3. Der Anteil der männlichen Fahrer unter den Motorradfahrern beträgt ca. 73% ✓ (0.5P)

4. Modus ✓mit x_{mod} =PKW✓ (0.5P+0.5P)

5. a) $f(MR|M) = \frac{0.11}{0.7}$✓$\approx 15.71\% \not> 30\%$ $\implies$ falsch✓ (1P)

   b) $f(M \cap PKW) + f(M \cap LKW) = 0.55 + 0.04 = 0.59$✓ $\implies$ richtig✓ (1P)

   c) $f(F|LKW) = \frac{0.01}{0.05} = 20\%$✓ $f(LKW|M) = \frac{0.04}{0.7} = 0.0571$✓
   $\implies$ richtig (1P)

6. $P(M) = P(M|MR) \cdot P(MR) + P(M|PKW) \cdot P(PKW) + P(M|LKW) \cdot P(LKW)$✓$=$
   $0.71 \cdot 0.08 + 0.5 \cdot 0.87 + 0.64 \cdot 0.05 = 0.5238$✓ (0.5P+0.5P)

7. a) $P(X_M \geq 1|Y_{MR} = 2, Y_{LKW} = 3) = 1 - P(X_M = 0|Y_{MR} = 2, Y_{LKW} = 3)$ (1P)
   $= 1 - \binom{2}{0} \cdot 0.71^0 \cdot 0.29^2 \cdot \binom{3}{0} \cdot 0.64^0 \cdot 0.36^3 = 1 - 0.0039 = 0.9961$✓✓

   b) $P(X_F \leq 2|Y_{PKW} = 5) = P(X_F = 0|Y_{PKW} = 5) + P(X_F = 1|Y_{PKW} = 5)$ (1P)
   $+P(X_F = 2|Y_{PKW} = 5) = \binom{5}{0}0.5^5 + \binom{5}{1}0.5^5 + \binom{5}{2}0.5^5 = 0.5$✓✓

Aufgabe 2

1. $x_5 = 176.5 * 8 - (176 + 181 + \ldots + 169) = \checkmark$

   $= 1412 - 1228 = 184 \checkmark$ (1P)

2. $r_{MF} = -0.2648 \checkmark$ (0.5P)

3. $$\begin{aligned} S^2_{M-F} &= \frac{1}{n-1}\sum_{i=1}^{n}\left((m_i - \overline{m}_n) - \left(f_i - \overline{f}_n\right)\right)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\Big((m_i - \overline{m}_n)^2 + \left(f_i - \overline{f}_n\right)^2 - 2\left(m_i - \overline{m}_n\right)\left(f_i - \overline{f}_n\right)\Big) \\ &= \frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)^2}{n-1} + \frac{\sum_{i=1}^{n}\left(f_i - \overline{f}_n\right)^2}{n-1} - 2\frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)\left(f_i - \overline{f}_n\right)}{n-1} \end{aligned}$$

   (1P)

4. a) $S^2_D = 97.6964 \checkmark$ (0.5P)

   b) $t_n = 1.8031 \checkmark$ (0.5P)

   c) $q^{0.9}_{t(7)} = 1.415 \checkmark$

   $6.375 - 1.415\frac{\sqrt{100}}{\sqrt{8}} = 1.3722. \checkmark$ (1P)

   d) $0 \notin [1.23,\ \infty)\ \Rightarrow$ Nullhypothese ablehnen$\checkmark$ (0.5P)

   e) $1 \notin [1.23,\ \infty)\ \Rightarrow$ Nullhypothese ablehnen.$\checkmark$ (0.5P)

   f) kürzer $\checkmark$ (0.5P)

5. a) Die Verteilung von D scheint rechtsschief $\checkmark$zu sein. (0.5P)

   b) Symmetrie $\checkmark$ (0.5P)

6. a) Binomialverteilung $\checkmark$mit $n = 8$ $\checkmark$und $p = 0.5$ $\checkmark$ (1.5P)

   b) 6 $\checkmark$ (0.5P)

7. a) 0.1094 $\checkmark$ (0.5P)

   b) 0.1445 $\checkmark$ (0.5P)

Aufgabe 3

1. $\lambda_Z = 2.2,\ \lambda_1 = 0.5,\ \lambda_2 = \lambda_Z - \lambda_1 = 1.7$ ✓ (0.5P)

2. $P(X_2 = 0; \lambda_2 = 1.0) = 0.3679$ ✓ (0.5P)

3. a) $P(Z > 2; \lambda_Z = 2.5) = 1 - P(Z \leq 2; \lambda_Z = 2.5)$ ✓
   $= 1 - 0.5438 = 0.4562$ ✓ (1P)

   b) $P(1 \leq X_1 \leq 4; \lambda_1 = 1.5)$ ✓
   $= 0.9814 - 0.2231 = 0.7582$ ✓ (1P)

4. a) $\tau = \left| \sqrt{10} \cdot \frac{2.5-3}{\sqrt{3}} \right| = 0.9128$ ✓ (0.5P)

   b) Verwerfe H_0, falls $0.9128 > \lambda_{1-\alpha/2} = 1.6448$ ✓
   $\Rightarrow\ H_0$ nicht ablehnen.✓ (1P)

   c) $pw = 2 \cdot P(\tau > 1.48)$ ✓
   $= 2 \cdot (1 - P(\tau \leq 1.48)) = 0.1388$ ✓ (1P)

   d) $pw \not< 0.1 \Rightarrow\ H_0$ nicht ablehnen.✓ (0.5P)

   e) H_0 gerade nicht ablehnen für $\alpha = 0.1388$ (alternativ: $\alpha = 0.13$).✓ (0.5P)

5. a) $f_X(x; p) = p \cdot (1-p)^x,\ x \geq 0,$
   $LL(p; x_1, ..., x_n) = ln(\prod_{i=0}^{n} p \cdot (1-p)^{x_i}) = ln\left(p^n (1-p)^{\sum x_i}\right)$
   $= n \cdot \log(p) + \log(1-p) \cdot \sum_{i=0}^{n} x_i$ ✓ (0.5P)

   b) $\frac{\partial LL}{\partial p} = \frac{n}{p} - \frac{\sum_{i=0}^{n} x_i}{1-p} = 0$ ✓
   $\Leftrightarrow \widehat{p}_{ML} = (1 + \overline{x}_n)^{-1}$ ✓ (1P)

   c) Hier: $\widehat{p}_{ML} = (1 + 5.5)^{-1} = 0.1538$ ✓
   $f_X(X = 3; p = \widehat{p}_{ML}) = 0.1538 \cdot (1 - 0.1538)^3 = 0.0932$ ✓
   (oder $f_X(X = 3; p = 0.17) = 0.0972$) (1P)

   d) $F_X(x; p) = 1 - (1-p)^x,\ x \geq 0,$
   $P(2 \leq X < 7; \widehat{p}_{ML} = 0.1538) = F_X(6) - F_X(1)$ ✓
   $= 0.4791$ ✓ (1P)
   (oder $P(2 \leq X < 7; p = 0.17) = 0.5031$)

Aufgabe 4

1. Komparatives Merkmal (0.5P)

2. median(*Größe*) = *groß* (0.5P + 0.5P)

3. Indifferenztabelle:

		Lokalität			
		X	Y	Z	Σ
Fußballfan	A	$n^*_{11} = 2.5$	$n^*_{12} = 0.8333$	$n^*_{13} = 1.6667$	5
	B	$n^*_{21} = 3.5$	$n^*_{22} = 1.1667$	$n^*_{23} = 2.3333$	7
	Σ	6	2	4	12

(2P)

4. $\phi^2 = 0.1944$ (1.5P)

5. $0 \leq \phi^2 \leq \min\{k-1, l-1\}$ entspricht $0 \leq \phi^2 \leq 1$ (0.5P)

6.
$$V = \sqrt{\frac{\phi^2}{\min\{k-1, l-1\}}} = 0.7325$$
(0.5P)

7. $0 \leq V \leq 1$ (0.5P)

8. Beide den Wert 0 (0.5P + 0.5P)

9. $n\phi^2 > q_{\chi^2(2),0.9} = 4.61$ (1P)

10. p-Wert 0.04 (1P)

11. p-Wert nicht kleiner als Irrtumswahrscheinlichkeit, H_0 nicht ablehnen. (0.5P)

Klausur Statistik (10 ECTS)

Aufgaben und Lösung

Name		**Prüfer**	Prof. Dr. I. Klein
Vorname		**Arbeitszeit**	Mittwoch, 30.07.2014 14:00 – 16:00 Uhr
Matrikelnummer			
Studienrichtung		**Sitzplatznummer**	
Semesterzahl		**Raum**	
Email (optional)			

Hinweise: Aufgabenblätter nicht auseinandertrennen!

Ergebnis:

Statistik	
Aufgabe	**Punkte**
1	
2	
3	
4	
Summe	
Note:	

Unterschrift des Kandidaten: ______________________________

Unterschrift des Prüfers: __________________________________

Hilfsmittel: Zugelassen sind eine selbsterstellte maximal vierseitige (DIN A4 doppelseitig) Formelsammlung, die vom Lehrstuhl zum Download bereitgestellte Tabellensammlung ohne Eintragungen und die R-Reference-Card (by Jonathan Baron) ohne Eintragungen.
Darüber hinaus sind Taschenrechner zugelassen. Es dürfen jedoch keine Programme oder Programmteile verwendet werden, die nicht fest in den Taschenrechner eingebaut sind. Alle Hilfsmittel sind selbst mitzubringen.

Bewertung: Für jede Aufgabe werden maximal zehn Punkte vergeben. Bewertet werden grundsätzlich nur Lösungen, die im Lösungsteil stehen und für die folgendes beachtet wird:

- Der Lösungsweg muss aus einer Darstellung der einzelnen Rechenschritte ersichtlich sein.
- Antworten sind stets zu begründen, es sei denn es wird ausdrücklich keine Begründung verlangt.
- Unleserliche Aufgabenteile werden mit 0 Punkten bewertet.

Viel Erfolg!

Aufgabe 1

Bei einer Polizeikontrolle werden die erfassten Fahrzeuge unter anderem auf die Merkmale „Fahrzeugtyp" (Motorrad(MR), PKW oder LKW) und „Geschlecht des Fahrzeugführers" (Männlich(M) oder Weiblich(F)) untersucht.

1. Um welchen Merkmalstyp handelt es sich beim Merkmal „Fahrzeugtyp"? Welche Skalierung ist für diesen Merkmalstyp sinnvoll?

2. Vervollständigen Sie die Tabelle der relativen Häufigkeiten ($f(\cdot)$).

	MR	PKW	LKW	Σ
M	0.11		0.04	0.7
F			0.01	
Σ	0.15	0.8		1

3. Interpretieren Sie die relative Häufigkeit $f(M|MR) \approx 0.73$.

4. Nennen und bestimmen Sie ein geeignetes Lagemaß für das Merkmal „Fahrzeugtyp".

5. Nehmen Sie auf Grundlage der Tabelle aus Teilaufgabe 2. zu folgenden Aussagen Stellung und begründen Sie diese:

   a) Mehr als 30% der kontrollierten männlichen Fahrer fahren Motorrad.

   b) Weniger als 60% der kontrollierten Fahrer sind männlich und fahren PKW oder LKW.

   c) Der Anteil der Frauen unter den kontrollierten LKW-Fahrern ist größer als der Anteil der LKW-Fahrer unter den kontrollierten Männern.

Laut Kraftfahrtbundesamt[1] waren in Deutschland 2014 71% der Motorradfahrer, 50% der PKW-Fahrer und 64% der LKW-Fahrer <u>männlich</u>.
Unter der Annahme, dass es nur die drei Fahrzeugtypen „Motorrad" (MR), „PKW" und „LKW" gibt, sind außerdem folgende Wahrscheinlichkeiten bekannt:

$P(MR) = 8\%$ $\qquad$ $P(PKW) = 87\%$

[1] http://www.kba.de/DE/Statistik/Kraftfahrer/Fahrerlaubnisse/Fahrerlaubnisbestand/fahrerlaubnisbestand_node.html

6. Berechnen Sie die Wahrscheinlichkeit, dass ein Fahrer männlich ist, d.h. $P(M)$.

Seien

X_M : „Anzahl der männlichen Fahrer"

X_F : „Anzahl der weiblichen Fahrer"

sowie

Y_i : „Anzahl der Fahrer eines Fahrzeugs vom Typ i"

mit $i \in \{$„PKW", „MR", „LKW"$\}$

7. Berechnen Sie die Wahrscheinlichkeit, dass

   a) unter 2 Motorrad-Fahrern und 3 LKW-Fahrern mindestens ein Fahrer männlich ist, d.h. $P(X_M \geq 1 | Y_{MR} = 2, Y_{LKW} = 3)$,

   b) unter 5 PKW-Fahrern höchstens zwei Fahrer weiblich sind, d.h. $P(X_F \leq 2 | Y_{PKW} = 5)$.

Schmierpapier zu Aufgabe 1

Aufgabe 2

Es wurde die Körpergröße von 8 zufällig ausgewählten Ehepaaren erhoben und in der folgenden Tabelle aufgeschrieben.

Ehepaar i	Körpergröße m_i des Ehemanns (in cm)	Körpergröße f_i der Ehefrau (in cm)
1	176	174
2	181	160
3	171	170
4	168	173
5		165
6	181	168
7	182	180
8	169	171

Mit den vorliegenden Daten soll die Vermutung getestet werden, dass Ehemänner im Mittel größer als ihre Ehefrauen sind. Fassen Sie die Einträge in obiger Tabelle als Realisationen einer Zufallsstichprobe mit den Zufallsvariablen

$$M_i : \text{„Körpergröße des } i\text{-ten Ehemanns"} \quad \text{und} \quad F_i : \text{„Körpergröße der } i\text{-ten Ehefrau"}$$

auf.
Realisierte Statistiken, die Sie bitte im Folgenden verwenden, sind $\overline{m}_8 = 176.5000$, $\overline{f}_8 = 170.1250$, $s_M^2 = 40.8571$, $s_F^2 = 36.4107$, $s_{MF} = -10.2143$.

1. Bestimmen Sie die Größe des fünften Ehemanns (m_5) in cm.

2. Berechnen Sie den Stichprobenkorrelationskoeffizienten r_{MF}.

3. Die Stichprobenvarianz von $D = M - F$ wird gegeben durch

$$s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left((m_i - f_i) - \left(\overline{m}_n - \overline{f}_n\right)\right)^2 .$$

Zeigen Sie folgende Gleichheit:

$$s_D^2 = \frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)^2}{n-1} + \frac{\sum_{i=1}^{n}(f_i - \overline{f}_n)^2}{n-1} - 2\frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)(f_i - \overline{f}_n)}{n-1}.$$

4. Führen Sie einen Mittelwertdifferenzentest für verbundene normalverteilte Stichproben zum Niveau $1 - \alpha = 0.9$ durch. Das Hypothesenpaar lautet:

$$H_0: \ \mu_M \leq \mu_F \quad \text{gegen} \quad H_1: \ \mu_M > \mu_F$$

a) Berechnen Sie für $D = M - F$ die Stichprobenvarianz s_D^2.

Verwenden Sie im Folgenden $s_D^2 = 100$, auch wenn Sie in Teilaufgabe 4a) einen anderen Wert herausbekommen haben.

b) Berechnen Sie die realisierte Prüfgröße t_n.

c) Geben Sie die realisierte Untergrenze des einseitigen Konfidenzintervalls an, das dem Test zugehörig ist.

Verwenden Sie im Folgenden die Untergrenze 1.23**, auch wenn Sie in Teilaufgabe 4c) einen anderen Wert herausbekommen haben**.

d) Treffen Sie die Testentscheidung basierend auf dem Konfidenzintervall.

e) Wie lautet die Testentscheidung, wenn das Hypothesenpaar zu

$$H_0: \ \mu_M \leq \mu_F + 1 \quad \text{gegen} \quad H_1: \ \mu_M > \mu_F + 1$$

geändert wird?

f) Wird das Niveau geändert, ändert sich die Länge eines Konfidenzintervalls. Wird es länger oder kürzer, wenn α vergrößert wird?

5. Der Schätzwert für das dritte standardisierte Moment der Verteilung von D beträgt $\mu_3 = 0.4511$.

a) Interpretieren Sie diesen Wert.

b) Wenn M und F bivariat normalverteilt sind, ist $D = M - F$ ebenfalls normalverteilt. Welches Schiefeverhalten weisen Normalverteilungen auf?

6. Es wird ein weiteres Vorgehen vorgeschlagen: Stammen die Größen von Männer und Frauen alle aus derselben Grundgesamtheit und finden Paare unabhängig von der Körpergröße zueinander, dann sind die Zufallsvariablen

$$G_i = \begin{cases} 1 & \text{falls der } i\text{-te Ehemanns größer ist, als seine Frau} \\ 0 & \text{falls der } i\text{-te Ehemanns nicht größer ist, als seine Frau} \end{cases}$$

jeweils Bernoulli-verteilt mit Erfolgswahrscheinlichkeit $p = 0.5$. Darüber hinaus sind diese Zufallsvariablen voneinander unabhängig.

a) Geben Sie die vollständig spezifizierte Verteilung der Zufallsvariablen $Z = \sum_{i=1}^{8} G_i$ an.

b) Berechnen Sie die Realisation von Z. Gehen Sie davon aus, dass $m_5 > f_5$, **unabhängig von den Berechnungen in Teilaufgabe 1**.

7. a) Eine faire Münze wird 8 Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass diese genau 2 Mal Kopf zeigt?

b) Eine faire Münze wird 8 Mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass diese höchstens 2 Mal Kopf zeigt?

Schmierpapier zu Aufgabe 2

Aufgabe 3

Laut offizieller FIFA-Statistik erzielte die deutsche Fußballnationalmannschaft in den letzten 10 Spielen durchschnittlich 2.2 Tore pro Spiel, wobei im Mittel 0.5 Tore in der ersten Hälfte des Spiels erzielt wurden. Im Folgenden seien

X_i: „Anzahl der in Halbzeit i geschossenen Tore"

mit $i = 1, 2$. Weiter sei angenommen, dass die X_i unabhängig poissonverteilt mit Parameter λ_i für $i = 1, 2$ ist.
Dann ist

Z: „Anzahl der geschossenen Tore in einem Spiel"

mit $Z = X_1 + X_2$ ebenfalls poissonverteilt mit Parameter $\lambda_Z = \lambda_1 + \lambda_2$.

1. Geben Sie die geschätzten Parameter λ_1, λ_2 sowie λ_Z auf Grundlage der letzten 10 Spiele an.

Verwenden Sie für alle nachfolgenden Teilaufgaben $\lambda_1 = 1.5$, $\lambda_2 = 1$ sowie $\lambda_Z = 2.5$

2. Berechnen Sie die Wahrscheinlichkeit, dass Deutschland in der zweiten Halbzeit des nächsten Länderspiels kein Tor erzielt?

3. a) Wie hoch ist die Wahrscheinlichkeit, dass Deutschland im nächsten Spiel mehr als zwei Tore erzielt?

b) Wie hoch ist die Wahrscheinlichkeit, dass Deutschland in der ersten Halbzeit mindestens ein Tor, aber nicht mehr als vier Tore schießt?

4.) Mittels eines approximativen Tests soll

$$H_0 : \lambda_Z = 3 \text{ gegen } H_A : \lambda_Z \neq 3$$

auf einem Signifikanzniveau von 10% getestet werden, d.h. die Teststatistik $T_n = \sqrt{n}\frac{\bar{Z}_n - \lambda_0}{\sqrt{\lambda_0}}$ wird als approximativ standardnormalverteilt angenommen.

a) Berechnen Sie die zugehörige Teststatistik.
Beachten Sie, dass ein konsistenter Schätzer für den Parameter λ_Z durch $\overline{Z}_n$ gegeben ist. Verwenden Sie im Folgenden $\overline{z}_{10} = 2.5$

b) Geben Sie den kritischen Bereich des Tests sowie Ihre Testentscheidung an.

c) Berechnen Sie für den approximativen Test den p-Wert. Gehen Sie hierbei von einem realisierten Wert der Teststatistik von 1.48 aus (**auch wenn Sie in Teilaufgabe 4a) einen anderen Wert herausbekommen haben**).

Sollten Sie aus Teilaufgabe 4c) kein Ergebnis vorliegen haben, so verwenden Sie im Folgenden einen p-Wert von 0.13

d) Fällen Sie Ihre Testentscheidung auf Basis des p-Werts aus Teilaufgabe 4c).

e) Geben Sie das größtmögliche Signifikanzniveau an, für welches Sie in Teilaufgabe 4d) die Nullhypothese nicht ablehnen können?

5. Die deutsche Fußballnationalmannschaft wurde in der Vergangenheit oft für ihre schlechte Chancenauswertung kritisiert. Die geometrische Verteilung kann z.B. dazu herangezogen werden, um die Anzahl der vergebenen Torchancen bis zum ersten Tor zu modellieren.

   Die Wahrscheinlichkeitsmassenfunktion f_X, und die kumulative Verteilungsfunktion F_X der geometrischen Verteilung lauten:

$$f_X(x;p) = p \cdot (1-p)^x \ , \tag{1}$$

$$F_X(x;p) = 1 - (1-p)^x \ , \tag{2}$$

   für $x = 0, 1, 2...$, wobei der Parameter p die Wahrscheinlichkeit eines Erfolgs misst.

a) Zeigen Sie, dass die Loglikelihoodfunktion der geometrischen Verteilung folgende Form besitzt:

$$LL(p; x_1, ..., x_n) = n \cdot ln(p) + ln(1-p) \cdot \sum_{i=0}^{n} x_i.$$

b) Leiten Sie den Maximum-Likelihood Schätzer für den Parameter p der geometrischen Verteilung her und geben Sie den Schätzwert an.

Ihnen sei nun bekannt, dass die deutsche Mannschaft in der Vergangenheit durchschnittlich 5.5 Torversuche pro Spiel benötigte bis zum ersten Mal ein Tor erzielt wurde. *(Sollte Ihnen aus Teilaufgabe 5b) kein Ergebnis vorliegen, verwenden Sie nachfolgend* $\widehat{p}_{ML} = 0.17$*)*

c) Wie hoch ist dann die geschätzte Wahrscheinlichkeit, dass beim vierten Torschuss das erste Tor erzielt wird?

d) Wie hoch ist die geschätzte Wahrscheinlichkeit, dass mindestens zwei aber weniger als sieben Torschüsse bis zum ersten Tor benötigt werden?

Schmierpapier zu Aufgabe 3

Aufgabe 4

Eine Universität hat vor 2 Jahren einen Masterstudiengang eingeführt, der nun bezüglich Zulassungskriterien und Studienerfolg analysiert werden soll. Von allen Studienanfängern des ersten Jahrgangs liegen im `R`-Dataframe `Master` folgende Merkmale vor.

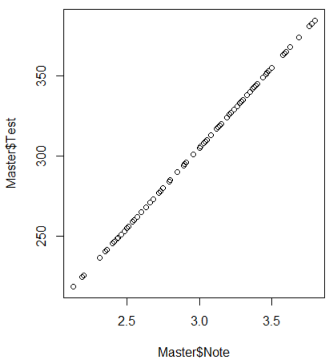
Merkmal *Note*:	Abschlussnote des Bachelor-Studiums	Spalte `Note`
Merkmal *Test*:	Erreichte Punktzahl im Zulassungstest	Spalte `Test`
Merkmal *Status*:	Aktueller Stand des Studiums	Spalte `Status`

Das Merkmal *Status* hat die Ausprägungen `abgebrochen`, `abgeschlossen` und `laufend`.

Zur Verfügung stehen Ihnen der zugehörige `R`-Output:

```
> abbruch=Master[Master$Status=="abgebrochen","Test"]
> (s1=sum(Master$Status=="abgebrochen"))
[1] 26
> (s2=sum(Master[Master$Status=="abgebrochen","Note"]))
[1] 77
> s2/length(abbruch)
[1] 2.961538
> quant=function(x){quantile(x,0.25,type=1) }
> (lage1=quant(abbruch))
25%
290
> (lage2=mean(abbruch))
[1] 301.1538
> (lage3=max(abbruch))
[1] 355
> (lage4=min(abbruch))
[1] 245
> plot(Master$Note,Master$Test)
```

und die mit dem letzten Befehl erstellte Abbildung:



1. Geben Sie einen geeigneten Datentyp für die Spalte `Status` an.

2. Geben Sie einen `R`-Befehl an, um die durchschnittliche Punktzahl der Studienanfänger im Zulassungstest zu berechnen.

3. Vervollständigen Sie die folgenden Aussagen:
   (a) Bisher haben ________ Studenten das Studium abgebrochen.
   (b) Die durchschnittliche Bachelornote der Studienabbrecher beträgt ________ .
   (c) 25% der Studienabbrecher hatten im Zulassungstest maximal ________ Punkte.

4. Sie interessieren sich für den Korrelationskoeffizienten der Merkmale *Note* und *Test*.
   (a) Schätzen Sie anhand der vorliegenden Informationen ab, welchen Wert dieser annimmt. (kurze Begründung)
   (b) Geben Sie einen `R`-Befehl an, um ihn zu berechnen.

5. Vervollständigen Sie die folgende `R`-Funktion `bar` so, dass damit ein Säulendiagramm absoluter Häufigkeiten von Merkmalsausprägungen in einem Datenvektor `z` erstellt wird.

```
bar=function(            ){
  haeuf=                 (z)
  barplot(                 )
}
```

6. Welche Kennzahl wird mit dem Befehl

```
> lage3-lage4
```

berechnet?

Es soll angenommen werden, dass die Zufallsvariable

$$X\text{: „logarithmierte Punktzahl beim Zulassungstest“}$$

in der Grundgesamtheit normalverteilt ist, d.h. $X \sim N(\mu_X, \sigma_X^2)$, und dass es sich bei den vorliegenden Daten um eine einfache Zufallsstichprobe handelt.

7. Bei einer Irrtumswahrscheinlichkeit von 10% soll getestet werden:

$$H_0 : \sigma_X^2 \geq \sigma_0^2 = 0.03 \qquad H_1 : \sigma_X^2 < \sigma_0^2$$

Führen Sie ausgehend von folgendem `R`-Code

```
> stich=log(Master$Test)
> k=length(stich)-1
```

den Test durch, indem Sie die `R`-Befehle angeben, um

(a) die Stichprobenvarianz als Objekt `s` zu berechnen,

(b) ausgehend von `s` die realisierte Prüfgröße als Objekt `p` zu berechnen,

(c) die kritische Schranke des Tests als Objekt `b` zu berechnen.

(d) Bei Verwendung der korrekten Befehle erhalten Sie bei Aufruf der Objekte

```
> s
[1] 0.0216875
> p
[1] 60.72501
> 1-p
[1] -59.72501
> b
[1] 67.87608
> pt(s,k)
[1] 0.5086257
```

Treffen Sie die Testentscheidung und begründen Sie diese anhand der relevanten Werte.

Schmierpapier zu Aufgabe 4

Lösung Klausur SS14 (10ECTS)

Aufgabe 1

1. qualitatives Merkmal (ungeordnet kategorial)✓, Nominalskala✓ (0.5P+0.5P)

2.

	MR	PKW	LKW	Σ
M	0.11	0.55	0.04	0.7
F	0.04	0.25	0.01	0.3
Σ	0.15	0.8	0.05	1

✓✓+ Rechenweg ✓ (1P+0.5P)

3. Der Anteil der männlichen Fahrer unter den Motorradfahrern beträgt ca. 73% ✓ (0.5P)

4. Modus ✓mit x_{mod} =PKW✓ (0.5P+0.5P)

5. a) $f(MR|M) = \frac{0.11}{0.7}$✓$\approx 15.71\% \not> 30\% \quad \Longrightarrow$ falsch✓ (1P)

   b) $f(M \cap PKW) + f(M \cap LKW) = 0.55 + 0.04 = 0.59$✓ $\Longrightarrow$ richtig✓ (1P)

   c) $f(F|LKW) = \frac{0.01}{0.05} = 20\%$✓ $f(LKW|M) = \frac{0.04}{0.7} = 0.0571$✓
   $\Longrightarrow$ richtig (1P)

6. $P(M) = P(M|MR) \cdot P(MR) + P(M|PKW) \cdot P(PKW) + P(M|LKW) \cdot P(LKW)$✓$=$
$0.71 \cdot 0.08 + 0.5 \cdot 0.87 + 0.64 \cdot 0.05 = 0.5238$✓ (0.5P+0.5P)

7. a) $P(X_M \geq 1|Y_{MR} = 2, Y_{LKW} = 3) = 1 - P(X_M = 0|Y_{MR} = 2, Y_{LKW} = 3)$ (1P)
   $= 1 - \binom{2}{0} \cdot 0.71^0 \cdot 0.29^2 \cdot \binom{3}{0} \cdot 0.64^0 \cdot 0.36^3 = 1 - 0.0039 = 0.9961$✓✓

   b) $P(X_F \leq 2|Y_{PKW} = 5) = P(X_F = 0|Y_{PKW} = 5) + P(X_F = 1|Y_{PKW} = 5)$ (1P)
   $+P(X_F = 2|Y_{PKW} = 5) = \binom{5}{0}0.5^5 + \binom{5}{1}0.5^5 + \binom{5}{2}0.5^5 = 0.5$✓✓

Aufgabe 2

1. $$x_5 = 176.5 * 8 - (176 + 181 + \ldots + 169) = \checkmark$$
   $$= 1412 - 1228 = 184 \checkmark$$ (1P)

2. $r_{MF} = -0.2648$ ✓ (0.5P)

3. $$\begin{aligned} S^2_{M-F} &= \frac{1}{n-1}\sum_{i=1}^{n}\left((m_i - \overline{m}_n) - \left(f_i - \overline{f}_n\right)\right)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\Big((m_i - \overline{m}_n)^2 + \left(f_i - \overline{f}_n\right)^2 - 2\,(m_i - \overline{m}_n)\left(f_i - \overline{f}_n\right)\Big) \\ &= \frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)^2}{n-1} + \frac{\sum_{i=1}^{n}\left(f_i - \overline{f}_n\right)^2}{n-1} - 2\frac{\sum_{i=1}^{n}(m_i - \overline{m}_n)\left(f_i - \overline{f}_n\right)}{n-1} \end{aligned}$$
   (1P)

4. a) $S^2_D = 97.6964$ ✓ (0.5P)
   b) $t_n = 1.8031$ ✓ (0.5P)
   c) $q^{0.9}_{t(7)} = 1.415$ ✓
      $6.375 - 1.415\frac{\sqrt{100}}{\sqrt{8}} = 1.3722.$ ✓ (1P)
   d) $0 \notin [1.23,\ \infty)$ $\Rightarrow$ Nullhypothese ablehnen✓ (0.5P)
   e) $1 \notin [1.23,\ \infty)$ $\Rightarrow$ Nullhypothese ablehnen.✓ (0.5P)
   f) kürzer ✓ (0.5P)

5. a) Die Verteilung von D scheint rechtsschief ✓zu sein. (0.5P)
   b) Symmetrie ✓ (0.5P)

6. a) Binomialverteilung ✓mit $n = 8$ ✓und $p = 0.5$ ✓ (1.5P)
   b) 6 ✓ (0.5P)

7. a) 0.1094 ✓ (0.5P)
   b) 0.1445 ✓ (0.5P)

Aufgabe 3

1. $\lambda_Z = 2.2,\ \lambda_1 = 0.5,\ \lambda_2 = \lambda_Z - \lambda_1 = 1.7$ ✓ (0.5P)

2. $P(X_2 = 0; \lambda_2 = 1.0) = 0.3679$ ✓ (0.5P)

3. a) $P(Z > 2; \lambda_Z = 2.5) = 1 - P(Z \leq 2; \lambda_Z = 2.5)$ ✓
   $= 1 - 0.5438 = 0.4562$ ✓ (1P)

   b) $P(1 \leq X_1 \leq 4; \lambda_1 = 1.5)$ ✓
   $= 0.9814 - 0.2231 = 0.7582$ ✓ (1P)

4. a) $\tau = \left| \sqrt{10} \cdot \frac{2.5-3}{\sqrt{3}} \right| = 0.9128$ ✓ (0.5P)

   b) Verwerfe H_0, falls $0.9128 > \lambda_{1-\alpha/2} = 1.6448$ ✓
   $\Rightarrow\ H_0$ nicht ablehnen.✓ (1P)

   c) $pw = 2 \cdot P(\tau > 1.48)$ ✓
   $= 2 \cdot (1 - P(\tau \leq 1.48)) = 0.1388$ ✓ (1P)

   d) $pw \nless 0.1 \Rightarrow\ H_0$ nicht ablehnen.✓ (0.5P)

   e) H_0 gerade nicht ablehnen für $\alpha = 0.1388$ (alternativ: $\alpha = 0.13$).✓ (0.5P)

5. a) $f_X(x; p) = p \cdot (1-p)^x,\ x \geq 0,$
   $LL(p; x_1, ..., x_n) = ln(\prod_{i=0}^n p \cdot (1-p)^{x_i}) = ln\left(p^n (1-p)^{\sum x_i}\right)$
   $= n \cdot \log(p) + \log(1-p) \cdot \sum_{i=0}^n x_i$ ✓ (0.5P)

   b) $\frac{\partial LL}{\partial p} = \frac{n}{p} - \frac{\sum_{i=0}^n x_i}{1-p} = 0$ ✓
   $\Leftrightarrow \widehat{p}_{ML} = (1 + \overline{x}_n)^{-1}$ ✓ (1P)

   c) Hier: $\widehat{p}_{ML} = (1 + 5.5)^{-1} = 0.1538$ ✓
   $f_X(X = 3; p = \widehat{p}_{ML}) = 0.1538 \cdot (1 - 0.1538)^3 = 0.0932$ ✓
   (oder $f_X(X = 3; p = 0.17) = 0.0972$) (1P)

   d) $F_X(x; p) = 1 - (1-p)^x,\ x \geq 0,$
   $P(2 \leq X < 7; \widehat{p}_{ML} = 0.1538) = F_X(6) - F_X(1)$ ✓
   $= 0.4791$ ✓ (1P)
   (oder $P(2 \leq X < 7; p = 0.17) = 0.5031$)

Aufgabe 4

1. factor (oder character) (0.5)

2. z.B. `mean(Master$Test)` (1.0)

3. (a) 26 (0.5)
   (b) 2.9615 (0.5)
   (c) 290 (0.5)

4. (a) Abbildung zeigt fast perfekt linearen Zusammenhang zwischen Note und Testpunktzahl. Somit sollte Korrelationskoeffizient Wert 1 oder einen Wert knapp darunter annehmen. (1.0)
   (b) `cor(Master$Note,Master$Test)` (1.0)

5. (1.5)

```
bar=function(z){
  haeuf= table(z)
  barplot( haeuf )
}
```

6. Spannweite des Merkmals Test bei den Studienabbrechern (1.0)

7. (a) `s=var(stich)` (0.5)
   (b) `p=k*s/0.03` (0.5)
   (c) `b=qchisq(0.1,k)` (0.5)
   (d) Da realisierte Prüfgröße mit 60.7 kleiner als kritische Schranke von 67.87, muss H_0 bei $\alpha = 0.1$ abgelehnt werden. (1.0)